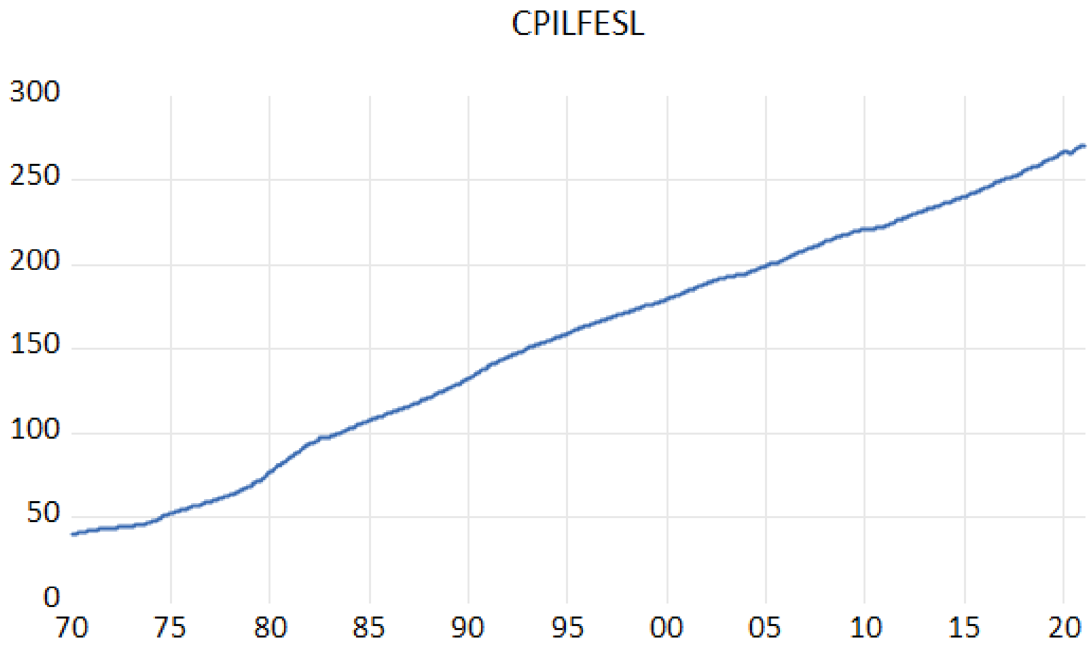
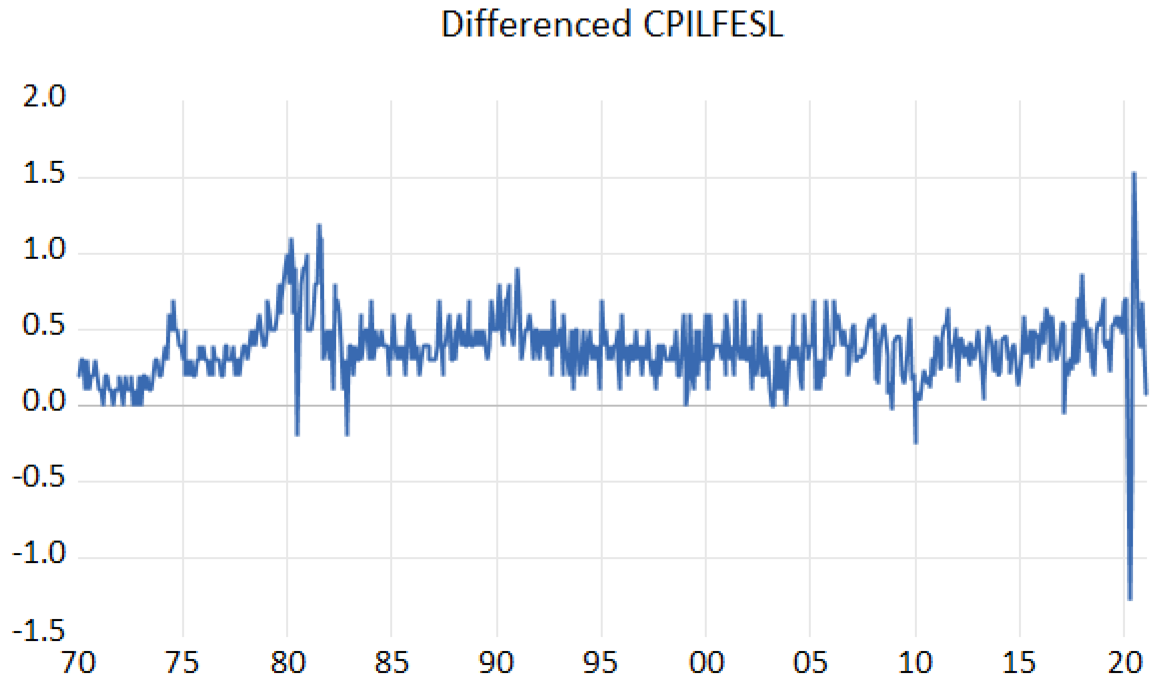
Assignment 3 – Paolo Sebastiani

1)





From the graph of the variable in levels, we can clearly see that it presents an upward trend which is quite linear, therefore it’s not I(0).
By taking the first difference, the series now looks much more stationary, even if the mean it’s not exactly zero and the variance is not always constant; however, for now we can conclude that it’s I(1), and then we’ll test it.

2)

Levels

Sample: 1970M01 2021M01
Included observations: 613

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.995	0.995	610.38	0.000
		2	0.991	-0.005	1216.1	0.000
		3	0.986	-0.003	1817.2	0.000
		4	0.982	-0.001	2413.7	0.000
		5	0.977	-0.004	3005.6	0.000
		6	0.972	-0.002	3592.8	0.000
		7	0.968	-0.001	4175.5	0.000
		8	0.963	0.002	4753.7	0.000
		9	0.959	-0.003	5327.3	0.000
		10	0.954	-0.006	5896.4	0.000
		11	0.949	-0.010	6460.8	0.000
		12	0.945	-0.006	7020.6	0.000
		13	0.940	-0.002	7575.7	0.000
		14	0.935	-0.002	8126.2	0.000
		15	0.931	-0.004	8672.0	0.000
		16	0.926	-0.002	9213.3	0.000
		17	0.921	-0.002	9749.9	0.000
		18	0.916	-0.003	10282.	0.000
		19	0.912	-0.003	10809.	0.000
		20	0.907	-0.003	11332.	0.000
		21	0.902	-0.004	11850.	0.000
		22	0.897	-0.004	12364.	0.000
		23	0.893	-0.004	12873.	0.000
		24	0.888	-0.004	13377.	0.000
		25	0.883	-0.003	13877.	0.000
		26	0.878	-0.002	14373.	0.000
		27	0.873	-0.003	14863.	0.000
		28	0.869	-0.003	15349.	0.000
		29	0.864	-0.003	15831.	0.000
		30	0.859	-0.003	16308.	0.000
		31	0.854	-0.005	16781.	0.000
		32	0.849	-0.004	17249.	0.000
		33	0.844	-0.005	17712.	0.000
		34	0.840	-0.003	18171.	0.000
		35	0.835	-0.004	18626.	0.000
		36	0.830	-0.003	19076.	0.000

First differences

Sample (adjusted): 1970M02 2021M01
Included observations: 612 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.472	0.472	136.97	0.000
		2	0.310	0.113	196.35	0.000
		3	0.176	-0.009	215.56	0.000
		4	0.156	0.065	230.54	0.000
		5	0.230	0.165	263.24	0.000
		6	0.253	0.099	302.86	0.000
		7	0.231	0.040	335.93	0.000
		8	0.288	0.159	387.41	0.000
		9	0.301	0.123	443.91	0.000
		10	0.259	0.024	485.67	0.000
		11	0.235	0.044	520.32	0.000
		12	0.180	0.007	540.72	0.000
		13	0.202	0.058	566.22	0.000
		14	0.147	-0.063	579.79	0.000
		15	0.163	0.019	596.58	0.000
		16	0.127	-0.040	606.79	0.000
		17	0.157	0.021	622.44	0.000
		18	0.149	-0.018	636.50	0.000
		19	0.173	0.034	655.44	0.000
		20	0.185	0.048	677.19	0.000
		21	0.152	-0.017	691.92	0.000
		22	0.120	-0.017	701.14	0.000
		23	0.085	-0.022	705.74	0.000
		24	0.061	-0.031	708.12	0.000
		25	0.050	-0.036	709.74	0.000
		26	0.067	-0.010	712.66	0.000
		27	0.043	-0.043	713.84	0.000
		28	0.069	-0.005	716.95	0.000
		29	0.045	-0.033	718.27	0.000
		30	0.054	-0.002	720.18	0.000
		31	0.037	-0.013	721.04	0.000
		32	0.011	-0.032	721.12	0.000
		33	0.036	0.041	721.94	0.000
		34	0.075	0.071	725.58	0.000
		35	0.020	-0.053	725.85	0.000
		36	-0.019	-0.048	726.09	0.000

Second differences

Sample (adjusted): 1970M03 2021M01
Included observations: 611 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.345	-0.345	73.080	0.000
		2	-0.023	-0.161	73.404	0.000
		3	-0.112	-0.206	81.120	0.000
		4	-0.088	-0.259	85.871	0.000
		5	0.053	-0.158	87.579	0.000
		6	0.049	-0.078	89.039	0.000
		7	-0.087	-0.198	93.698	0.000
		8	0.034	-0.156	94.425	0.000

In the correlogram of the variable in levels, the ACF presents a very slow decay, confirming that the series is not I(0). Passing to the first differences, instead, we see a quick decay after 1 or 2 lags, even if there are some following lags that shows value not very close to zero. However, if we analyze the correlogram of the second differences, then it’s clear that there is the problem of “over-difference”, since an artificial negative correlation is introduced.

In conclusion, we can state that the series is I(1).

3)

Null Hypothesis: CPILFESL has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 9 (Automatic - based on SIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.021077	0.5881
Test critical values:	1% level	-3.973346	
	5% level	-3.417289	
	10% level	-3.131040	

*MacKinnon (1996) one-sided p-values.

CPILFESL in levels: The DF test on the most general model (i.e. with both the intercept and the trend) suggest that there is a unit root

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CPILFESL)
Method: Least Squares
Date: 02/10/25 Time: 18:11
Sample (adjusted): 1970M11 2021M01
Included observations: 603 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CPILFESL(-1)	-0.003200	0.001583	-2.021077	0.0437
D(CPILFESL(-1))	0.360072	0.040735	8.839360	0.0000
D(CPILFESL(-2))	0.096923	0.043170	2.245168	0.0251
D(CPILFESL(-3))	-0.091009	0.043609	-2.086909	0.0373
D(CPILFESL(-4))	-0.027968	0.043483	-0.643198	0.5203
D(CPILFESL(-5))	0.137963	0.043152	3.197102	0.0015
D(CPILFESL(-6))	0.072257	0.043463	1.662499	0.0969
D(CPILFESL(-7))	-0.051211	0.045023	-1.137435	0.2558
D(CPILFESL(-8))	0.127644	0.044509	2.867848	0.0043
D(CPILFESL(-9))	0.158386	0.043256	3.661558	0.0003
C	0.195365	0.058162	3.358993	0.0008
@TREND("1970M01")	0.001243	0.000619	2.009237	0.0450

9 lags

Wald Test:
Equation: FULL_MODEL

Test Statistic	Value	df	Probability
F-statistic	2.048491	(2, 591)	0.1298
Chi-square	4.096981	2	0.1289

Null Hypothesis: C(2)=0, C(3)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	0.001245	0.000619
C(3)	-0.003204	0.001585

Restrictions are linear in coefficients.

2° step of DF test: The F-test on $\beta = \varphi = 0$ suggest that we can remove the coefficient β from the model (indeed the series does not show a quadratic trend)

Null Hypothesis: CPILFESL has a unit root
Exogenous: Constant
Lag Length: 9 (Automatic - based on SIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.238806	0.9308
Test critical values:	1% level	-3.440965	
	5% level	-2.866115	
	10% level	-2.569265	

*MacKinnon (1996) one-sided p-values.

3° step of DF test: Also in the model with only the constant, H_0 cannot be rejected

Wald Test:
Equation: ONLY_INTERCEPT

Test Statistic	Value	df	Probability
F-statistic	7.611509	(2, 592)	0.0005
Chi-square	15.22302	2	0.0005

Null Hypothesis: C(1)=0, C(2)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(1)	0.093144	0.028229
C(2)	-2.59E-05	0.000108

Restrictions are linear in coefficients.

4° step of DF test: The F-test on $\alpha = \varphi = 0$ suggest that we should maintain α in the model (as expected, since the series shows a clear linear trend).

Now we can go back to the output of the DF test on the model with only the constant, and by confronting the t-Statistic of φ with the critical values of the normal distribution, we cannot reject H_0 and thus conclude that the series has a unit root.

Null Hypothesis: D(CPILFESL) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 8 (Automatic - based on SIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.022720	0.0085
Test critical values:	1% level	-3.973346	
	5% level	-3.417289	
	10% level	-3.131040	

*MacKinnon (1996) one-sided p-values.

DF test on the differenced series: now the null hypothesis on the presence of a unit root is immediately rejected with a high significance, and therefore we conclude that the series is *I(1)*.

4)

	AIC	BIC
ARIMA (3,1,3)	-0.552687	-0.494952
ARIMA (3,1,4)	-0.552614	-0.487662
ARIMA (4,1,2)	-0.558335	-0.500600
ARIMA (4,1,4)	-0.549931	-0.477762
ARIMA (5,1,3)	-0.552786	-0.480618
ARIMA (5,1,4)	-0.562206	-0.482820
ARIMA (5,1,5)	-0.577630	-0.491028

I've selected 7 models which produced white noise residuals, and computed their AIC and BIC. As we can see, the model with the lowest AIC is not the same as the one with the lowest BIC. Therefore, since we're working with a quite large sample (>600 observations), we could use the BIC as the main criterion and thus select the ARIMA(4,1,2) model, which is more parsimonious than the ARIMA(5,1,5). On the other hand, the models with the lowest AIC tend to be the better one for forecasting (as we'll see in the next point).

5)

		RMSE	MAE
ARIMA(4,1,2)	Static	0.43323	0.258485
	Dynamic	0.499433	0.30254
ARIMA(5,1,5)	Static	0.410075	0.239171
	Dynamic	0.479745	0.290611

As expected, the model which forecasts the best among the two (by using both the *Root Mean Squared Error* and the *Mean Absolute Error* as criteria) is the ARIMA(5,1,5), both in static and dynamic forecasts.